Copy No._____

Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

ASSA Inversion Model Upgrades: Implementation of Raytrace AEL, and Bayesian Posterior Probability Density Distributions

Alex MacGillivray and Melanie Austin
JASCO Research Ltd

JASCO Research Ltd
Suite 2101, 4464 Markham St
Victoria, BC V8Z 7X8

Project Manager: Scott Carr, 250-483-3300

Contract Number: W7707-021923/001/HAL

Contract Scientific Authority: Francine Desharnais, 902-426-3100 ext 219

Defence R&D Canada – Atlantic

Contract Report
DRDC Atlantic CR 2003-251
January 2005

Canada

ASSA Inversion Model Upgrades: Implementation of Raytrace AEL, and Bayesian Posterior Probability Density Distributions

Alex MacGillivray, Melanie Austin
JASCO Research Ltd

JASCO Research Ltd
Suite 2101, 4464 Markham St
Victoria, BC
V8Z 7X8

Project Manager: Scott Carr, 250-483-3300

Contract number: W7707-021923/001/HAL

Contract Scientific Authority: Francine Desharnais, 902-426-3100 ext 219

Defence R&D Canada – Atlantic

Contract Report

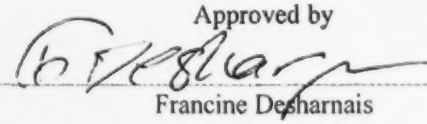
DRDC Atlantic CR 2003-251

January 2005

Author

Alex MacGillivray

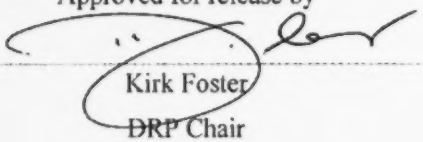
Approved by



Francine Desharnais

Scientific Authority

Approved for release by



Kirk Foster

DRP Chair

Terms of release: The scientific or technical validity of this Contract Report is entirely the responsibility of the contractor and the contents do not necessarily have the approval or endorsement of Defence R&D Canada.

Terms of release: The information contained herein is proprietary to Her Majesty and is provided to the recipient on the understanding that it will be used for information and evaluation purposes only. Any commercial use including use for manufacture is prohibited. Release to third parties of this publication or information contained herein is prohibited without the prior written consent of Defence R&D Canada.

© Her Majesty the Queen as represented by the Minister of National Defence, 2005

© Sa majesté la reine, représentée par le ministre de la Défense nationale, 2005

Abstract

The Adaptive Simplex Simulated Annealing (ASSA) Fortran code currently used by DRDC Atlantic to solve a series of geoacoustic inversion problems has been expanded to solve a new type of problem and to provide Bayesian statistical analysis of the model result uncertainties. The new problem type involves the localization of acoustic sources and receivers given relative arrival time data. A new forward model has been added to the program for solving the localization problem using a ray tracing approach. A new cost function, to be minimized in the optimization of the localization problem, has also been added to the program based on the root mean square (rms) time difference of the relative arrival times. The Bayesian analysis is performed using a Gibbs sampling approach to determine *a posteriori* probability density (PPD) distributions, providing estimates of model uncertainties for geoacoustic inversion problems. In order to preserve the general nature of the ASSA code and to allow future program enhancements, the new features have been added in a modular fashion and are pure extensions to the existing code that do not change the main functionality of the original ASSA program.

Résumé

Le code Fortran de recuit simulé simplex adaptatif (ASSA) qu'utilise actuellement RDDC Atlantique pour résoudre une série de problèmes d'inversion géoacoustique a été étendu pour solutionner un nouveau type de problème et pour permettre l'analyse statistique bayesienne des incertitudes dans les résultats du modèle. Le nouveau type de problème comporte la localisation des sources et des récepteurs acoustiques à partir de données de temps relatif d'arrivée. Un nouveau modèle direct a été ajouté au programme dans le but de résoudre le problème de localisation selon la technique du traçage de rayon. Une nouvelle fonction de coût, qui sera minimisée dans l'optimisation du problème de localisation, a également été ajoutée au programme et se fonde sur la différence quadratique moyenne des temps relatifs d'arrivée. L'analyse bayesienne, exécutée selon une technique d'échantillonnage de Gibbs, vise à déterminer les répartitions des densités de probabilité *a posteriori* (PPD) et fournit une évaluation des incertitudes du modèle pour des problèmes d'inversion géoacoustique. Afin de préserver la nature générale du code d'ASSA et de permettre les améliorations futures du programme, les nouvelles caractéristiques ont été ajoutées de façon modulaire et constituent de pures extensions du code existant, qui ne changent pas la fonction principale du programme ASSA original.

This page intentionally left blank.

106-04026/

Executive summary

Introduction

In shallow water, sonar performance is strongly affected by the acoustic characteristics of the seafloor. Knowledge of these characteristics can be used to improve sonar performance indirectly, by suggesting changes in the use of the sonar, or directly, through inclusion in the processing applied to the received signals. DRDC Atlantic currently has the ability to solve a series of geo-acoustic inversion problems using an Adaptive Simplex Simulated Annealing (ASSA) FORTRAN program. This versatile tool was expanded to look at new problems.

Results

A new optimization for the localization of acoustic sources and receivers was added to the ASSA suite. The existing ASSA code used the RAMGEO forward model in the inversion of geoacoustic parameters, however this model was not suitable for a passive localization problem. Thus, a new forward model was added to the code that uses a ray trace approach to solve for relative arrival times. A new cost function, based on the root mean square (RMS) time difference of the relative arrival times, has also been added to the existing ASSA code for the localization optimization.

A method to calculate posterior probability density distributions of the model parameters was also added to aid in determining the uncertainty of the geo-acoustic estimates obtained with ASSA. A Gibbs sampling approach has been employed to calculate estimates of the Bayesian posterior probability density distributions. While the implementation of the Bayesian analysis in the ASSA program has attempted to generalize the formulation, it should not yet be used with problems other than geo-acoustic inversion.

Military Relevance

With these new additions, the code can be used to solve a source/receiver localization problem. This feature can be used for Array Element Localization (AEL) to refine the positions of an array of receivers. It can also be used to localize passive acoustic sources detected on an array of receivers (i.e. marine mammal localization). These two problems are directly relevant to naval operations. In addition, the code can be used to advance knowledge in geo-acoustic inversion problems by looking at uncertainties in parameter estimates. Geo-acoustic inversions are becoming the standard in naval system support, as the environment is rarely sampled directly.

Future Work

The new additions to the ASSA inversion code with the additions of these two new problems remain to be fully tested. No new addition to the ASSA framework is planned at this time.

MacGillivray, A.O. and Austin, M.E. 2003. Inversion model upgrades. DRDC Atlantic CR 2003-251. Defence R&D Canada – Atlantic.

Sommaire

Introduction

En eaux peu profondes, le rendement d'un sonar dépend beaucoup des caractéristiques acoustiques du fond marin. La connaissance de ces caractéristiques permet d'améliorer le rendement d'un sonar de façon indirecte, en faisant ressortir des modifications dans l'utilisation du sonar, ou de façon directe, en permettant une inclusion dans le traitement appliqué aux signaux reçus. RDDC Atlantique peut actuellement solutionner une série de problèmes d'inversion géoacoustique au moyen d'un programme en Fortran de recuit simulé simplex adaptatif (ASSA). Cet outil polyvalent a été étendu à l'examen de nouveaux problèmes.

Résultats

Une nouvelle optimisation de la localisation des sources et des récepteurs acoustiques a été ajoutée au programme ASSA. Le code ASSA existant utilisait le modèle direct RAMGEO pour l'inversion des paramètres géoacoustiques, mais ce modèle ne convenait pas au problème de la localisation passive. Ainsi, un nouveau modèle direct ajouté au code utilise une technique de traçage de rayon pour résoudre les temps relatifs d'arrivée. Une nouvelle fonction de coût, fondée sur la différence quadratique moyenne des temps relatifs d'arrivée, a également été ajoutée au code ASSA existant pour l'optimisation de la localisation.

Une méthode servant à calculer les répartitions des densités de probabilité *a posteriori* (PPD) des paramètres du modèle aide en outre à déterminer l'incertitude des évaluations géoacoustiques obtenues avec ASSA. Une technique d'échantillonnage de Gibbs est employée pour évaluer les répartitions bayesiennes des densités de probabilité *a posteriori*. Bien que l'analyse bayesienne du programme ASSA tente de généraliser la formulation, il ne convient pas encore de l'appliquer à des problèmes autres que d'inversion géoacoustique.

Pertinence militaire

Grâce à ces nouveaux ajouts, le code peut servir à solutionner un problème de localisation de sources/récepteurs. Il peut être utilisé pour la localisation d'éléments de réseau (AEL), afin d'accroître la précision des positions dans un réseau de récepteurs. Il permet également de localiser des sources acoustiques passives détectées dans un réseau de récepteurs (p. ex. pour la localisation des mammifères marins). Ces deux problèmes offrent une pertinence directe en matière d'opérations navales. Par ailleurs, le code peut faire progresser la connaissance des problèmes d'inversion géoacoustique par un examen des incertitudes dans les évaluations des paramètres. Les inversions géoacoustiques sont de plus en plus courantes en matière de soutien des systèmes navals, car l'environnement fait rarement l'objet d'échantillonnages directs.

Plans futurs

Les nouvelles additions au code d'inversion ASSA pour ces deux nouveaux problèmes doivent encore être testées. Aucune addition future n'est prévue pour le logiciel intégré ASSA.

MacGillivray, A.O. and Austin, M.E. 2003. Inversion model upgrades. DRDC Atlantic CR 2003-251. Defence R&D Canada – Atlantic.

Table of contents

Abstract i

Executive summary iii

Sommaire v

Table of contents vii

List of figures ix

1. Acoustic Source and Receiver Localization Problem 1
    1.1 Downhill Simplex Method 1
    1.2 Modules added to the ASSA program 2
        1.2.1 Ray Trace Implementation 2
        1.2.2 RMS relative time difference cost function 2
    1.3 Summary of the modifications to the existing ASSA code 3
        1.3.1 Addition of the new problem 3
        1.3.2 Addition of the new model 5
        1.3.3 Addition of the new cost function 6
        1.3.4 Summary of new modules 6
    1.4 Testing of the new modules 7

2. The Bayesian Posterior Probability Density (PPD) Distribution Module 11
    2.1 The Gibbs sampler: computing the PPD 11
    2.2 Important statistical properties of the PPD 12
    2.3 The implementation of the Gibbs sampler 13
    2.4 Running the Bayesian PPD module 14
        2.4.1 Problem-specific parameters from ASSAIN.DAT employed by the Bayesian module 15
        2.4.2 Bayesian specific input section of ASSAIN.DAT 16
    2.5 Summary of output files 18
        2.5.1 COVM.DAT 18
        2.5.2 MPPD.DAT 19

2.5.3 MODELS.DAT 20
2.6 Summary of additions/modifications to the ASSA code 20

3. References 23

Annex A – Sample Input File for Localization Problem 24

Annex B – Sample Secondary Input File for Localization Problem 27

Annex C – Sample Input File for Bayesian Module 28

List of symbols/abbreviations/acronyms/initialisms 31

Distribution list 32

List of figures

Figure 1: Source positions (blue stars) and receiver positions for the synthetic test data solution. Blue crosses indicate 'true receiver positions' and red crosses indicate the final receiver positions calculated by the inversion. 8

Figure 2: Receiver positions. Blue crosses indicate 'true receiver positions', black crosses indicate randomly selected initial receiver positions and red crosses indicate the final receiver positions. 9

Figure 3: Convergence of the x coordinate of the receiver positions for a synthetic test case. 10

List of tables

Table 1 :Format of the problem specific data portion of ASSAIN.DAT 4

Table 2 :Format of the model specific data portion of ASSAIN.DAT 6

Table 3. Files required for the localization problem. 6

Table 4. New format for data line 5 of section 1 of ASSAIN.DAT (c.f., [4, p. 8]). 15

Table 5. Generic/problem-specific portions of ASSAIN.DAT that are employed by the Bayesian module. 16

Table 6. Format for section 6 (i.e., the Bayesian-specific section) of ASSAIN.DAT. 17

Table 7. Output format for COVM.DAT. 19

Table 8. Output format for MPPD.DAT. 20

Table 9. Output format for MODELS.DAT. 20

Table 10. Files comprising the Bayesian module. 21

Table 11. Summary of changes to the original ASSA code, pertaining to the Bayesian module. 22

1. Acoustic Source and Receiver Localization Problem

New modules have been added to the ASSA code to solve a generalized Array Element Localization (AEL) problem. The optimization problem assumes fixed distributed receivers and uses direct path relative arrival times to invert the three-dimensional positions of multiple sources and multiple receivers. Estimates of the receiver positions and the relative arrival times are given as inputs to the model. *A priori* estimates of the source and receiver positions, with their associated uncertainties, are also provided. The problem is solved using a downhill simplex approach and a ray tracing method is used as the forward model to compute the relative arrival times at each simplex step. Model convergence is determined based on the minimization of a cost function involving the RMS difference of the model derived and the measured relative arrival times.

1.1 Downhill Simplex Method

For the generalized localization problem considered in this contract, it is assumed that the fixed receiver positions are known approximately and that the relative arrival times are given. The optimization is carried out using the downhill simplex method with the objective of minimizing the rms time difference of the modeled and the true relative arrival times. Due to the nature of the problem the solution will rapidly converge to a local minimum and a global minimization approach is not necessary. The existing ASSA code first uses simulated annealing to approach the local minimum and then goes into a 'quenching' stage to get the final solution of the geoacoustic inversion problems. For the AEL problem the only part of the ASSA process that is required is the final quenching stage, in which a pure downhill simplex is performed.

Initially a random simplex of models is generated, subject to the constraint that the model parameters lie within the minimum and maximum limits defined in the ASSA input file, and the downhill simplex method is carried out. Setting the ASSA code to perform the pure downhill simplex directly is achieved by setting the NTEMP value from Line 3 of the main ASSA input file to zero. Setting NTEMP to zero forces the code to go directly to the 'quenching' process since the simulated annealing loop of the ASSA code is only executed while the variable ITEMP (initialized to 1) is less than or equal to NTEMP (refer to Line 231 in the main ASSA.f module of the code).

The 'quenching' process continues until the model converges or the maximum number of iterations, defined in the input file, has been reached. Convergence of the model parameters is obtained when the value of the cost function has been minimized to a sufficient degree. The convergence criteria value is a parameter that is given in Line 5 of the ASSA input file. For the problem under present consideration the convergence criteria value of 1E-5 was found to be a suitable value. Convergence for

the test case described below was obtained after approximately 15 mins of computation time on a 800 MHz computer.

1.2 Modules added to the ASSA program

1.2.1 Ray Trace Implementation

The modules that have been added to the ASSA program to perform the ray tracing are directly based on existing code written in the IDL programming language by Nicole Collison and Stan Dosso. The idl code was originally written to solve the problem of tracking a field of freely drifting sonobuoys, given relative arrival time measurements and approximations for source and initial sonobuoy positions. Although the minimization approach used for the sonobuoy problem is not related to the approach taken for this contract, the ray trace model used to solve for the direct path arrival times is the same. The theory behind the ray trace model is described in Dosso and Collison [5] and the reader is referred to their work for a more detailed discussion.

AEL inversions may be carried out using relative arrival times as the data to invert, requiring the source emission instants to be included in the inversion. Alternatively, differences between the relative arrival times can be used as the data to invert and the source emission instants are not necessary [6]. Dosso and Collison include the source emission instants as parameters to be inverted in the idl code developed for sonobuoy tracking. For the current implementation these parameters have been removed from the model and the arrival times are expressed relative to a reference time defined as the arrival time for the receiver that is best localized (i.e. the receiver for which the uncertainty associated with its position is the smallest). For receivers that have the same positional uncertainty, the one that occurs first in the list of model parameters is chosen as the reference. This allows the problem to be easily expanded to the situation where the receivers are not necessarily required to be fixed in space, but rather they are constrained to be fixed relative to one another.

1.2.2 RMS relative time difference cost function

The cost function that is minimized in the optimization problem for this contract is defined in Equation 1 :

$$\Phi = \sqrt{\frac{\sum_{i=1}^{nsrc}\sum_{j=1}^{nrec}(t_true_{ij} - t_m_{ij})^2}{nsrc * nrec}} \qquad (1.1)$$

where t_true is the measured data and t_m is the arrival time calculated with the forward model. The problem involves nsrc sources and nrec receivers, giving nsrc*nrec arrival time values. The arrival times are expressed relative to the reference time as described above and the RMS time difference is calculated. The value of the

cost function is referred to as the 'Energy' and is the value that is minimized in the downhill simplex.

1.3 Summary of the modifications to the existing ASSA code

A new problem, model and cost function have been added to the ASSA code following the instructions laid out in section 4 of the original ASSA Contractor Report [4]. The following sections of this report describe the additions and modifications that have been made to the original code for the addition of the new optimization problem.

1.3.1 Addition of the new problem

The existing ASSA code was able to solve four different geoacoustic inversion problems that are referred to throughout the code as Problems 1 through 4. The new localization problem that has been added for this contract will be referred to as Problem 5. The PMC_MATRIX.INC include file that is required by the ASSA code has been updated to reflect the addition of the new problem. Within this include file the NUM_PROBLEM constant has been increased to 5 and the PROB_MOD array has been expanded to include the new combinations allowable for the new problem and new forward model.

A new variable NSRC has been added to the PROBVARS.INC common area to hold the value of the number of acoustic sources for Problem 5 that is read from the ASSAIN.DAT input file. An array named T_TRUE has also been added to the PROBVARS.INC common area to hold the true arrival time data. Finally, variables for the a priori estimates of the source and receiver positions (x, y and z) and their positional uncertainties have been added to PROBVARS.INC. The names of these variables can be found in Table 1 below.

A new constant, NSRC_MAX has been added to the include file PARAMS.INC. Also the value of NDIM_MAX in this include file has been increased due to the fact that Problem 5 requires 3*NHYD + 3*NSRC (three spatial parameters for each source and for each hydrophone) model parameters to obtain a solution. Since the values of NHYD_MAX and NSRC_MAX are set to 40 the value of NDIM_MAX must be 240. Increasing the value of NDIM_MAX to satisfy this requirement resulted in memory problems with respect to the STRACK array which was being used to track the model parameter values throughout the optimization process. The dimensions of this STRACK array became too large and it was decided that the approach to tracking this data would be revised. Instead of storing the STRACK data into a large array as the program proceeds and then writing it to a file after final covergence is obtained, the STRACK values are now output directly to a file as the program proceeds. To achieve this modification the STRACK matrix was removed from the GENVARS.INC common area and all calls that updated the STRACK matrix (which occurred in SIMPLEX.f and in ASSA.f) were replaced with calls to output the STRACK data to a file named STRACKTEMP.DAT. Once convergence is obtained, a new file named STRACK.DAT is opened, the appropriate header information (i.e.

number of iterations, NMOD and NDIM) are output to the file and the contents of STRACKTEMP.DAT are copied into it. The file STRACKTEMP.DAT is deleted from memory at the end of the program.

The maximum and minimum limits of the model parameters are read from the READ_LIMITS.f module. This module has been updated to read in the limits for Problem 5 in the same manner as the limits are read in for Problem 4.

It should be noted that Line 3 of Section 1 of the input file must assign a value of 0 to the NTEMP parameter for the localization problem.

The problem specific data is read from the input file using a new module named READ_PROB_5. This new module follows the design of the existing modules for reading the problem specific data for the geoacoustic inversion problems and uses the same functions for reading data from a file and for error checking. The format of the problem specific section of the input file is shown in Table 1.

Table 1. *Format of the problem specific data portion of ASSAIN.DAT*

SEC.	DATA LINE	CONTENTS	DATA TYPE	DESCRIPTION
2	1	*PROB_DATA_FILE*	Character*125	The file name of the secondary input file.
	2	*NHYD*	Integer	Number of receivers. Defined in PROBVARS.INC.
	3	*RX0(i)*	Real	The a priori estimate of the x coordinate of receiver(i). Defined in PROBVARS.INC.
		RY0(i)	Real	The a priori estimate of the y coordinate of receiver(i). Defined in PROBVARS.INC.
		RZ0(i)	Real	The a priori estimate of the z coordinate of receiver(i). Defined in PROBVARS.INC.
		DRX0(i)	Real	Uncertainty of the a priori estimate of the x coordinate of receiver(i). Defined in PROBVARS.INC.
		DRY0(i)	Real	Uncertainty of the a priori estimate of the y coordinate of receiver(i). Defined in PROBVARS.INC.
		DRZ0(i)	Real	Uncertainty of the a priori estimate of the z coordinate of receiver(i). Defined in PROBVARS.INC.
	4	*NSRC*	Integer	Number of sources. Defined in PROBVARS.INC.
	5	*SX0(i)*	Real	The a priori estimate of the x coordinate of source(i). Defined in PROBVARS.INC.
		SY0(i)	Real	The a priori estimate of the y coordinate of source (I Defined in PROBVARS.INC.

		SZ0(i)	Real	The a priori estimate of the z coordinate of source (i). Defined in PROBVARS.INC.
		DSX0(i)	Real	Uncertainty of the a priori estimate of the x coordinate of source (i). Defined in PROBVARS.INC.
		DSY0(i)	Real	Uncertainty of the a priori estimate of the y coordinate of source (i). Defined in PROBVARS.INC.
		DSZ0(i)	Real	Uncertainty of the a priori estimate of the z coordinate of source (i). Defined in PROBVARS.INC.

The module READ_PROBLEM_DATA has also been updated to call READ_PROB_5 as necessary.

An additional input file is required for the localization problem that contains the measured arrival time data in seconds. The name of this file is read from the first line of the problem specific data section of ASSAIN.DAT. The variable holding the file name has been added to PROBVARS.INC. The first line of this input file must contain two values indicating the number of columns and number of rows in the remainder of the file. The remainder of the file is then read with arrival time data looping over each receiver and each source with the receiver index changing the fastest. An example of the secondary input file is shown in the Annex B.

1.3.2 Addition of the new model

The existing ASSA code implemented the RAMGEO forward model (referred to in the code as Model 1) to solve the geoacoustic inversion problems. A new ray tracing forward model has been added to the program to solve the localization problem and is referred to as Model 2. Two new modules, RAY_TRACE.f and RAY_D.f, have been added to the code and are used to implement the forward model. In accordance with the instructions in the original ASSA Contractor Report [4] the NUM_MODEL constant in the PMC_MATRIX.INC include file has been increased to 2 and the PROB_MOD and MOD_COST arrays have been updated accordingly.

The model requires a sound speed profile as input and the existing ZI and PROFI arrays defined in the MODVARS.INC common area are used for to hold these values. An array, T_MOD, used to hold the arrival times calculated by the forward model has also been added to MODVARS.INC.

A new module RUN_MODEL_2 has been added to the program following the design of the existing RUN_MODEL_1 module. In the new module the model parameters (three-dimensional receiver and source positions) are defined, the ray tracing modules

are called to calculate the arrival times for the current model parameters and then the new module which calculates the new cost function is called.

The existing READ_MODEL_DATA module has been updated to call a new module, READ_MODEL_2 that reads the ray trace model specific data from the input file in the same manner that the existing module READ_MODULE_1 reads in Model 1 specific data. The only data required by the model is the sound speed profile, which is read into the ZI (depth) and PROFI (sound speed) arrays. The format of the model specific data portion of the input file is given in Table 2.

Table 2. *Format of the model specific data portion of ASSAIN.DAT*

SEC.	DATA LINE	CONTENTS	DATA TYPE	DESCRIPTION
4	1	*ZI*	Real	Depths (m) for a sound speed profile read with its associated PROFI in pairs until the end of their values is signified by a ZI value being less than or equal to zero.
		PROFI	Real	Sound speeds (m/s) for a sound speed profile read with its associated ZI in pairs until the end of their values is signified by a ZI value being less than or equal to zero.

1.3.3 Addition of the new cost function

The new RMS time difference based cost function is calculated using a module named RMSRTD.f, which has been added to the software, and is referred to as COSTFN 2. As required, the NUM_COSTFN constant in the PMC_MATRIX.INC include file has been updated to 2 and the MOD_COST array has been updated accordingly. COSTFN 2 will be associated with Model 2 and is called from RUN_MODEL_2 as necessary. There is no cost function specific data that is required as input.

1.3.4 Summary of new modules

A list of the modules that have been added to the ASSA code and a brief description of their purpose is given in Table 3 below.

Table 3. *Files required for the localization problem.*

FILENAME	SUBROUTINE	FUNCTION
RAY_TRACE.f	RAY_TRACE	Solves the forward problem of computing direct ray arrival times. This module is called from RUN_MODEL_2.f. This module has been ported from the idl code written by Nicole Collison and Stan Dosso.
RAY_D.f	RAY_D	Traces direct rays for a given value of the ray parameter p. This module is called by RAY_TRACE.f. This module has been ported from the IDL code written by Nicole Collison and

		Stan Dosso.
sgemv.f	sgemv	This is a module from the BLAS library and is used in the cost function calculation. This module performs matrix –vector calculations.
sdot.f	sdot	This is a module from the BLAS library and is used in the cost function calculations. This module calculates a dot product.
xerbla.f	xerbla	This function is from the BLAS library and is called by sgemv.f
READ_PROB_5.f	READ_PROB_5	Reads in problem specific data from ASSAIN.DAT for the localization problem
READ_MODEL_2.f	READ_MODEL_2	Read in model specific data from ASSAIN.DAT for the ray trace model
RUN_MODEL_2.f	RUN_MODEL_2	Calls the ray trace modules and the cost function calculation modules.
RMSRTD.f	RMSRTD	Calculates the RMS relative time difference base cost function for the localization problem.

1.4 Testing of the new modules

To test the performance of the new forward model and cost function, a subset of the synthetic test data that was used by Dosso and Collison for the sonobuoy tracking problem was input to the ASSA code. For the sonobuoy problem, the receivers were free to drift and the inversion was used to solve for sonobuoy tracks. For the general localization problem considered for this contract the receivers are assumed to be fixed therefore only the initial sonobuoy positions were used for the test results shown in this report.

The test results are shown in Figures 1 and 2 below. The source positions (indicated by the blue stars in Figure 1) were distributed about the centers of the boundary defined by ±3000m in x and y, at a depth of 180m, as defined in the sonobuoy problem. The 'true' receiver positions (indicated with blue crosses in Figures 1 and 2) were placed at depth of 100m at x and y positions as defined for the sonobuoy problem. The 'true' arrival times were calculated by running the forward model portion of the IDL code written by Dosso and Collison and were put into the secondary, problem specific, input file for the ASSA program. The arctic sound speed profile that was used by Dosso and Collison was also used for this test.

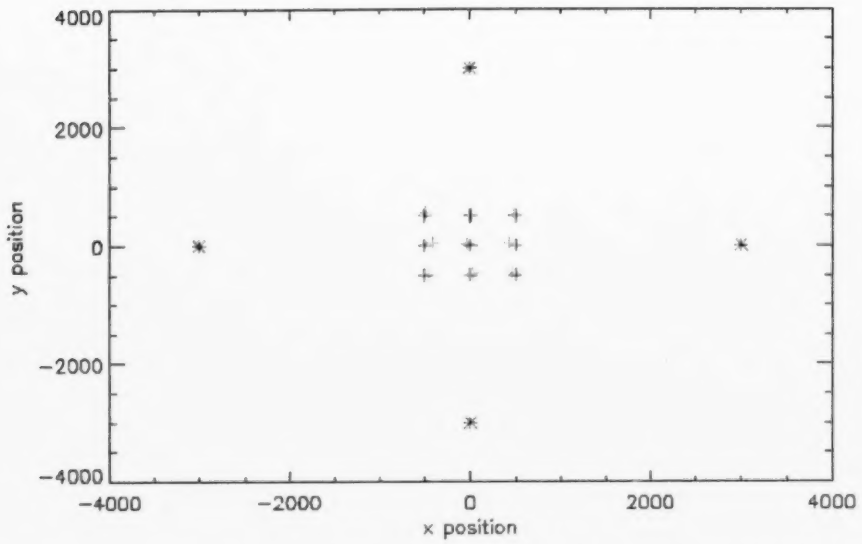


Figure 1: *Source positions (blue stars) and receiver positions for the synthetic test data solution. Blue crosses indicate 'true receiver positions' and red crosses indicate the final receiver positions calculated by the inversion.*

In order to check consistency of the results the test was performed numerous times. The results shown in Figures 1 and 2 are the average of three optimization runs. Figure 1 illustrates the synthetic true model plotted with the optimization results. The blue stars indicate the source positions, the blue crosses indicate the true receiver positions and the red crosses indicate the inverted receiver positions. The receiver positions are shown more clearly in Figure 2 where the blue crosses represent the 'true' receiver positions, the black crosses indicate the initial receiver positions at the beginning of the downhill simplex (randomly chosen to lie within the constraints defined in the ASSAIn input file) and the red crosses indicate the optimized receiver positions returned by the program. It can be seen that the inverted receiver positions represent the true data better than initial random receiver distribution does.

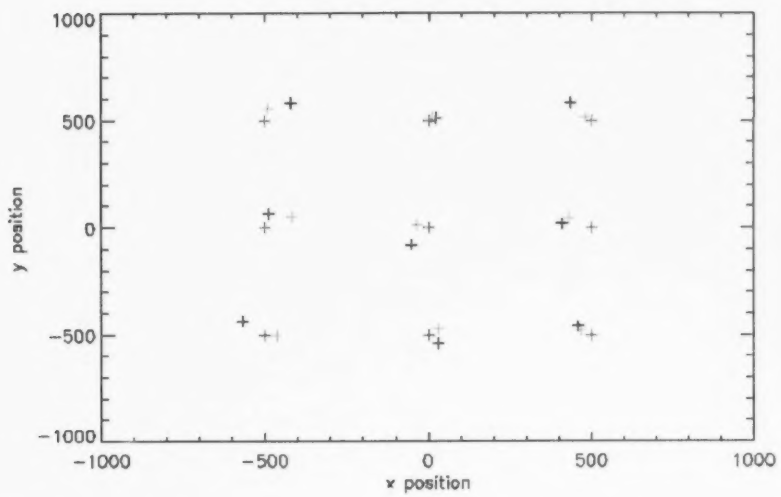


Figure 2: *Receiver positions. Blue crosses indicate 'true receiver positions', black crosses indicate randomly selected initial receiver positions and red crosses indicate the final receiver positions.*

The convergence of the model parameters during the downhill simplex process of one test run is illustrated in Figure 3. The first plot in the top left corner of Figure 3 shows the value of 2*(Emax-Emin)/(Emax+Emin) where Emax and Emin are the maximum and minimum values of the cost function at a given iteration. This value must reach the convergence criteria value (set to 1e-5 for the test cases) for convergence to be obtained. The remaining plots indicate the convergence of the x position of the receivers. Note that in the early steps of the downhill simplex, the x positions are randomly distributed and vary greatly, however as the cost function reaches minimization the x positions converge to a single value. As seen in Figures 1 and 2 the value to which the coordinates converge are close to the true model values. For the test case illustrated in Figure 3 convergence occurs after approximately 400 steps. It can also be seen that after approximately 200 steps most of the receiver coordinates have nearly converged to a single value and the convergence criteria value is beginning to level off.

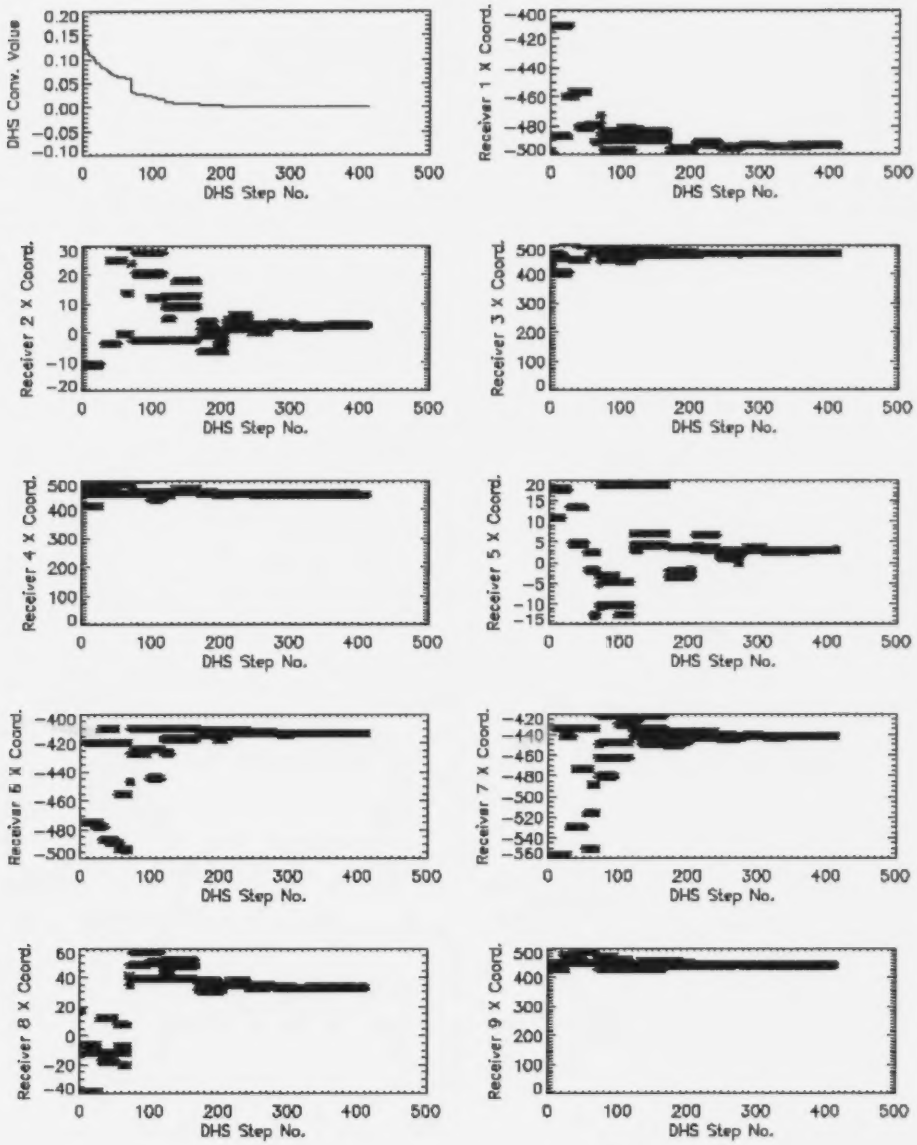


Figure 3: *Convergence of the x coordinate of the receiver positions for a synthetic test case.*

2. The Bayesian Posterior Probability Density (PPD) Distribution Module

For inverse problems, it is often desirable to associate uncertainties with the maximum-likelihood (ML) estimates of model parameters that have been derived by the inversion process. It is possible, using the techniques of Bayesian *a posteriori* statistics, to estimate model uncertainties by computing a posterior probability density (PPD) distribution for the parameters an inverse problem. Recently, a Bayesian formulation of matched-field inversion has been developed by Dosso [1], whereby the PPD is estimated from an unbiased Gibbs sampling of the parameter space.

For the present contract, JASCO has extended the functionality of ASSA, so that the program now computes posterior probability densities for the model parameters of a geoacoustic inversion problem. The implementation of the Gibbs sampler follows that of Gillard *et al.* [2] and the Bayesian module itself is based on Chris Gillard's PPDINVERSE code. Gillard's code has been totally re-written, however, with the goal of generalizing the Bayesian formulation, so that it is applicable to classes of inverse problems other than geoacoustic inversion.

2.1 The Gibbs sampler: computing the PPD

In the Bayesian formulation of inverse theory (as described in Dosso's article [1]) the posterior probability density (PPD) for an inverse problem is proportional to the product of the prior distribution, $P(\mathbf{m})$, and the likelihood function, $L(\mathbf{d}|\mathbf{m})$:

$$P(\mathbf{m} \mid \mathbf{d}) \propto L(\mathbf{d} \mid \mathbf{m}) P(\mathbf{m}) \qquad (2.1)$$

where $\mathbf{m}$ is the vector of model parameters, $\mathbf{d}$ is the data vector and $P(\mathbf{m}|\mathbf{d})$ is the PPD of the model parameters, given the data. Assuming unbiased Gaussian errors, the likelihood function may be expressed in terms of the error function, $E(\mathbf{m})$:

$$L(\mathbf{d} \mid \mathbf{m}) \propto \exp\left(-E(\mathbf{m})\right) \qquad (2.2)$$

In general, the form of $E(\mathbf{m})$ (also referred to as the Gibbs energy) is problem specific. For matched-field processing, Dosso suggests an error function of the form:

$$E(\mathbf{m}) = \sum_f B_f(\mathbf{m}) \left| \mathbf{d}_f \right|^2 / \nu_f \qquad (2.3)$$

where $B_f(\mathbf{m})$ is the Bartlett mismatch, $|\mathbf{d}_f|^2$ is the squared amplitude of the data vector and ν_f is the data variance at each frequency f. Note that, if the data vector is normalized (as is usually the case with MFP), then $|\mathbf{d}|^2 = 1$ and $|\mathbf{d}|^2/\nu$ is the signal-to-noise ratio (SNR).

If one assumes a uniform prior distribution (i.e., $P(\mathbf{m}) = \text{const}$), which is generally the case when the inversion parameters are unknown, then the normalized PPD may be written as:

$$P(\mathbf{m} \mid \mathbf{d}) = \frac{\exp(-E(\mathbf{m}))}{\int \exp(-E(\mathbf{m}'))d\mathbf{m}'} \quad (2.4)$$

Note that the denominator of (3) is constant and thus trivial to evaluate if the numerator is known, since the PPD is subject to the normalization condition $\int P(\mathbf{m} \mid \mathbf{d})\, d\mathbf{m} = 1$. An efficient technique for discretely evaluating equation (2.4), known as *Gibbs sampling*, is described by Dosso in a recent article [1]. The principle of Gibbs sampling is quite simple: the numerator of equation (2.4) is sampled over the entire parameter space, using the Metropolis test[1] at a system temperature of $T = 1$. Recall that, in the Metropolis algorithm, a particular set of model parameters, $\mathbf{m}$, is accepted or rejected with probability $P = \exp(-E(\mathbf{m})/T)$. The density of models accepted by the Metropolis test in a volume of parameter space $\delta\mathbf{m}$ is simply proportional to the value of the PPD in that region. For a more in-depth discussion of Gibbs sampling, and of Bayesian statistics in general, the reader is referred to Dosso's paper [1].

2.2 Important statistical properties of the PPD

Once the PPD of an inverse problem is known, then it is possible to compute the mean model, $\langle \mathbf{m} \rangle$, and covariance matrix, $\mathbf{C}_\mathbf{M}$, of the model parameters from the moments of PPD:

$$\langle \mathbf{m} \rangle = \int \mathbf{m}' P(\mathbf{m}' \mid \mathbf{d})\, d\mathbf{m}' \quad (2.5)$$

$$\mathbf{C}_\mathbf{M} = \int (\mathbf{m}' - \langle \mathbf{m} \rangle)(\mathbf{m}' - \langle \mathbf{m} \rangle)^T P(\mathbf{m}' \mid \mathbf{d})\, d\mathbf{m}' \quad (2.6).$$

The mean model is the vector of expectation values of the model parameters and, unless the marginal distributions of the parameters are symmetric (*e.g.*, Gaussian), the mean is generally different from the maximum likelihood model, $\hat{\mathbf{m}}$. Elements of the covariance matrix give the covariances between the various model parameters. Two particularly important quantities given by the covariance matrix are the model variance and cross-correlation. Note that the variances of the model parameters are simply given by the diagonal entries of the covariance matrix. In addition, given $\mathbf{C}_\mathbf{M}$, one may compute the model correlation matrix, which is defined by:

[1] Details of the Metropolis algorithm (also known as simulated annealing) will not be discussed here, as there already exists a great body of literature dealing with the subject.

$$R_{ij} = \frac{C_{M\,ij}}{\sqrt{C_{M\,ii} \cdot C_{M\,jj}}} \qquad (2.7)$$

Entries of the model correlation matrix, which are between [-1, 1], measure the degree of correlation between the different model parameters (see [1, p.136]).

Of particular interest, in addition to the model mean and covariance matrix, are the marginal probability distributions, P_k, which describe the one-dimensional distributions of the individual model parameters. The marginal distributions may be thought of as projections of the PPD onto the axes of the parameter space. Since the Gibbs sampling process actually computes the PPD discretely, the marginal distributions are approximated by normalized histograms of the Gibbs sampled model parameters.

2.3 The implementation of the Gibbs sampler

The new Bayesian module of ASSA implements the PPD computation according to the algorithm described by Gillard *et al.* [2] in their article to appear in IEEE JOE, and implemented by Gillard in his PPDINVERSE Fortran code. In order to improve the efficiency of the Gibbs sampling (*i.e.*, to reduce the number of rejected models), Gillard *et al.* employ three important optimizations, which have been incorporated into the ASSA Bayesian module:

1. The Gibbs sampler is initiated from the maximum likelihood model, $\hat{\mathbf{m}}$. This avoids a high rejection rate early on in the algorithm.

2. Model perturbations are performed in a rotated parameter space, defined by the principal axes of the model covariance matrix $\mathbf{C_M}$. This results in improved sampling efficiency, as models are more likely to be perturbed along "valleys" of the cost-function in the rotated parameter space.

3. The estimate of $\mathbf{C_M}$, (and thus the rotation matrix) is continually updated as the Gibbs sampling proceeds. In this way, the algorithm reaps the benefits of sampling in a rotated parameter space before the covariance matrix has strictly converged.

The Gibbs sampler is an iterative algorithm, and therefore requires some convergence criterion to determine when the correct answer has been achieved. The convergence test used in the Bayesian PPD module is the same as that employed by both Dosso [1] and Gillard [2]: two separate Gibbs samplers are run, in parallel, to compute two independent estimates of the PPD and the algorithm is assumed to have converged when the difference between the two PPD estimates is within some user-specified tolerance. At present, the difference between the two PPD estimates is defined to be the maximum absolute difference between the independent estimates of the (discrete) cumulative marginal distributions:

$$\varepsilon = \max_{j,k} \sum_{i=1}^{j} \left| \begin{matrix} P_k^{(1)}\left((i - \tfrac{1}{2})\Delta m_k + m_k^{(\min)}\right) \\ -P_k^{(2)}\left((i - \tfrac{1}{2})\Delta m_k + m_k^{(\min)}\right) \end{matrix} \right| \Delta m_k \quad (2.8)$$

where m_k is the k^{th} model parameter, $P_k^{(1)}$ and $P_k^{(2)}$ are the two independently estimated marginal distributions and Δm_k is the bin-width. When $\varepsilon \leq \varepsilon_{\text{conv}}$, the PPD has converged and the final PPD is computed from the union of the two independent estimates of the PPD.

Another feature, optionally employed by the Bayesian PPD module, is to estimate the data variance, v, from the maximum-likelihood model. Examining equation 2.3, one should note that the data variance must be known in order to compute the Gibbs energy for a given model, and thus for computing the PPD. However, for inversion problems using real data, the data variance is often unknown and must be estimated from the data itself. For matched-field problems, Dosso and Nielsen [3] proposed a method for estimating the data variance from the maximum-likelihood model. It can be shown that the data variance may be estimated by:

$$\hat{v} = B(\hat{\mathbf{m}})|\mathbf{d}|^2 / N \quad (2.9)$$

where $B(\hat{\mathbf{m}})$ is the Bartlett mismatch at the maximum-likelihood model, $|\mathbf{d}|^2$ is the squared amplitude of the data vector and N is the number of statistically uncorrelated sensors (which is equivalent to the number of propagating modes, for matched-field problems). The reader is referred to Dosso and Nielsen's article [3] for a more in depth discussion of this topic.

Another feature, that may or may not be used for computing the PPD, is the cascaded Metropolis test, as employed by Gillard *et al.* [2]. For multi-frequency inversion, the Metropolis test is typically performed only once for each model, using the frequency-averaged value of the cost function. In the cascaded test, the Metropolis test is performed at each frequency in-turn, and the model is rejected if any of the individual tests fail. Using the cascaded Metropolis test may speed up convergence of the PPD estimation, but it is unknown whether results computed with the cascaded test compare favourably with those computed using the frequency-averaged Metropolis test. Thus, the Bayesian PPD module allows the user to select which Metropolis test is to be used.

2.4 Running the Bayesian PPD module

The Bayesian PPD module is a new addition to the functionality of ASSA, and thus employs a number of extensions to the main input file, ASSAIN.DAT. All of these extensions, save one, appear at the end of the input file, in order to maintain backwards compatibility as much as possible. The old format of ASSAIN.DAT is described in the original ASSA contractor report [4, § 3.2.1], and the reader is referred to that document for the details of running the original ASSA code.

The first step required for running the Bayesian PPD module of ASSA is to set the boolean flag *DO_PPD* on data line 5 of section 1 (generic data) of ASSAIN.DAT to ".TRUE.". This parameter has been appended to data line 5 of ASSAIN.DAT and *must* be present in order for the input file to parse correctly. The new form of data line 5 is presented in Table 4. It should be noted that, if one wishes to execute an old ASSAIN.DAT file with the new version of ASSA, the only change required is to append the value ".FALSE." to the end of data line 5, to inhibit execution of the Bayesian module.

Table 4. *New format for data line 5 of section 1 of ASSAIN.DAT (c.f., [4, p. 8]).*

DATA LINE	PARAMETER #	CONTENTS	DATA TYPE	DESCRIPTION
5	1	*DO_GRID*	Logical	Ignored.
	2	*DO_ASSA*	Logical	If .TRUE. then ASSA analysis is performed. If .FALSE. then ASSA analysis is skipped (see *DO_PPD*).
	3–5	*TRACK, SCREEN_PR, Q_RAN*	Logical	As defined in the original ASSA program (see [4, p. 8] for details).
	6	*DO_PPD*	Logical	If .TRUE. then Bayesian PPD analysis is performed. If *DO_PPD* is .TRUE. and *DO_ASSA* is .FALSE. then PPD analysis is performed without first conducting ASSA minimization. In this case, the value of the maximum-likelihood model *must* be given on the TRUE_DATA line of section 3 of ASSAIN.DAT (see [4, p. 10]).

2.4.1 Problem-specific parameters from ASSAIN.DAT employed by the Bayesian module

In addition to the Bayesian-specific parameters, which are input at the end of ASSAIN.DAT (as discussed in the next section), there are a number of other parameters, both generic and problem-specific, which are employed by the Bayesian PPD module. These parameters specify the form of the cost function (used for computing the Gibbs energy), the dimensionality of the inversion problem and the bounds of the parameter space. Note that the format of these parameters has not been modified in any way and that use of these parameters in no way affects the execution of the main ASSA code. The names of these parameters, and details of how they are employed by the Bayesian PPD module, are outlined in Table 5.

Table 5. *Generic/problem-specific portions of ASSAIN.DAT that are employed by the Bayesian module.*

SEC.	DATA LINE	PARAM. #	CONTENTS	DATA TYPE	DESCRIPTION
1	2	1–3	*PROBLEM, MODEL, COSTFN*	Integer	These codes determine the type of inversion problem, model and cost function used for evaluating the Gibbs energy for a given set of model parameters. Currently, the output of the Bayesian module is only valid for PROBLEM=1–4, MODEL=1 and COSTFN=1 (*i.e.*, Geoacoustic inversion).
	3	7	*NDIM*	Integer	The dimensionality of the parameter space of the inversion problem. That is, the number of model parameters for which the PPD is to be computed.
		8	*TRUE_DATA*	Logical	Only used if *DO_ASSA* is .FALSE. and *DO_PPD* is .TRUE.; in this case, *TRUE_DATA* must be equal to .TRUE. and the maximum likelihood model must be specified on data line 4 of section 3 of ASSAIN.DAT.
2	1	1	*NUMEAVE*	Integer	The number of frequencies at which the cost function is to be evaluated. If the user has selected the cascaded Metropolis test, then the test is evaluated at each frequency. If the user has selected the classic Metropolis test, then the test is only evaluated once, for the average Gibbs energy over all frequencies.
3	2	1–*NDIM*	*MINLIM(i)*	Real	The lower bounds of the parameter space searched by the Gibbs sampler. This is also the lower bounds for the marginal parameter distributions computed by the Bayesian PPD module.
	3	1–*NDIM*	*MAXLIM(i)*	Real	The upper bounds of the parameter space searched by the Gibbs sampler. This is also the upper bounds for the marginal parameter distributions computed by the Bayesian PPD module.
	4	1–*NDIM*	*TRUE(i)*	Real	If *DO_ASSA* is .FALSE. and *DO_PPD* is .TRUE. then this is assumed to contain the maximum likelihood model. Thus, the Gibbs sampling is initiated from this model and the data variance may be estimated from the value of the cost function at this model.

2.4.2 Bayesian specific input section of ASSAIN.DAT

If the *DO_PPD* flag is set to .TRUE., then an additional section must appear at the end of ASSAIN.DAT, (*i.e.*, after the cost-function specific data) specifying parameters

necessary for execution of the Bayesian PPD module. Table shows the input format for the Bayesian-specific section of ASSAIN.DAT, and provides a description of each of the input parameters.

There are some subtle points to note, regarding the input parameters described in Table 6:

1. Since there are really two Gibbs samplers running in parallel, convergence checks are actually performed each time 2 × *NCHECK* Gibbs samples are accumulated. Likewise, for the termination condition, execution actually terminates if 2 × *MAXSAMP* samples are accumulated without achieving convergence.

2. If the user desires to specify the data variance, v_f, directly, it should be understood that the Bayesian module assumes a normalized data vector, $|\mathbf{d}_f|^2=1$; thus, the variances specified on data line 2 are actually values of $v_f/|\mathbf{d}_f|^2$.

3. If the data variance, v_f, is to be estimated from the maximum-likelihood model then the number of statistically uncorrelated sensors at each frequency, N_f, must be estimated independently by the user and entered on data line 2. In general, for matched-field problems, N_f is equal to the number of propagating modes. As well, Dosso has indicated that N_f may also be estimated by principal component analysis of the data covariance matrix; the reader is referred to his article for details [1, pp. 148–149].

4. Since the marginal distributions are used to test for convergence, lower *NBINS* values (*i.e.*, coarser sampling of the marginal distributions) result in speedier convergence at the expense of accuracy.

Table 6. *Format for section 6 (i.e., the Bayesian-specific section) of ASSAIN.DAT.*

DATA LINE	PARAMETER #	CONTENTS	DATA TYPE	DESCRIPTION
1	1	*EPSILON*	Real	The convergence criterion, ε_{conv}, for computing the PPD (see equation 2.8); must be in the range (0, 1). Once convergence has been achieved, the Bayesian PPD module outputs its results and terminates.
	2	*MAXSAMP*	Integer	The maximum number of Gibbs samples to accumulate, without achieving convergence. If the number of Gibbs samples exceeds *MAXSAMP*, then the Bayesian PPD module terminates with an error message indicating that convergence was not achieved.
	3	*CASCADE*	Logical	A boolean flag specifying whether to use the cascaded Metropolis test (*CASCADE*=.TRUE.) or the frequency averaged Metropolis test (*CASCADE*=.FALSE.).

	5	*NCHECK*	Integer	The number of Gibbs samples to accumulate between convergence checks. As there is very little computational overheard associated with performing a convergence check, this should be a number on the order of *NDIM*.
	6	*ESTVAR*	Logical	A boolean flag indicating whether the data variance should be estimated from the maximum likelihood model (*ESTVAR*=.TRUE.) or specified directly by the user (*ESTVAR*=.FALSE.). This flag affects the type of data input on the next data line.
2	1–*NUMEAVE*	*N(i)* or *V(i)*	Real	**If *ESTVAR*=.TRUE.** then this is the number of statistically uncorrelated sensors at each frequency, N_f. This value is used to estimate the data variance (see equation 2.9). **If *ESTVAR*=.FALSE.** then this is the normalized data variance at each frequency, $v_f/\|\mathbf{d}_f\|^2$. This is also equal to the inverse of the signal-to-noise power ratio, 1/SNR.
3	1–*NDIM*	*NBINS(i)*	Integer	The number of histogram bins, between *MINLIM*(i) and *MAXLIM*(i), for computing the marginal distribution for each model parameter.

Note: Data lines 2 and 3 may be safely broken across multiple lines of input. However, parameter values from data lines 2 and 3 may *not* be mixed on the same input line (*i.e.*, at least one line-break must separate data line 2 from data line 3).

2.5 Summary of output files

2.5.1 COVM.DAT

The output file COVM.DAT contains the mean model, $\langle \mathbf{m} \rangle$, and covariance matrix, $\mathbf{C}_\mathbf{M}$, as estimated by the Bayesian PPD module. The format for the COVM.DAT data file is given in Table 7. One should note that the values in COVM.DAT are updated every time *NDIM* parameter perturbations are performed. Thus, if the file can be viewed in read-only mode, during program execution, it may be possible to monitor the progress of the PPD estimation by examining the contents of COVM.DAT[2].

[2] Depending on the operating system on which ASSA is run, opening output files in read-only mode while the program is running may or may not interfere with program execution. Opening output files in read-only mode has been observed to work on both the Windows and Unix platforms.

Table 7. Output format for COVM.DAT.			
LINE(S)	**PARAMETERS**	**LINE FORMAT**	**DESCRIPTION**
1	*NDIM*	(I6)	The number of elements of the mean model vector.
2	*MEAN(NDIM)*	(*NDIM* G15)	The vector of expectation values (*i.e.*, the mean model).
3	*NDIM, NDIM*	(I6, X, I6)	The number of rows and columns of the covariance matrix.
4...*NDIM* + 4	*COVM(NDIM, NDIM)*	(*NDIM* G15)	The covariance matrix.

2.5.2 MPPD.DAT

The output file MPPD.DAT contains the normalized marginal posterior probability densities, for each of the model parameters, as computed by the Bayesian PPD module. The format for the MPPD.DAT data file is given in Table 8. Note that, since the two Gibbs samplers produce two independent estimates of the PPD, the marginal distributions in this file are actually computed from the union of the two PPD estimates. The values in MPPD.DAT are updated each time a convergence check is performed. Thus, it may be possible to monitor the progress of the PPD estimation by viewing the contents of MPPD.DAT as the program runs.

The file MPPD.DAT contains the values of the marginal distributions in the matrix *MPPD*, with $2\times$*NDIM* columns and *MAXNBINS* rows. The abscissa (bin centre) and ordinate (marginal probability) values of the marginal distributions are interleaved in the columns of *MPPD*. Thus, the x-values of the marginal distribution of parameter k are contained in column $(2k-1)$ of MPPD and the y-values are contained in column $2k$. If the value of *NBINS* for parameter k is less than *MAXNBINS*, then only rows $1\rightarrow$*NBINS*(k) contain marginal distribution values; the remaining rows are padded with zeros. Note that the marginal distributions have been normalized so that:

$$\int P_k(m_k)\,dm_k \cong \sum_{i=1}^{NBINS(k)} P_k\left((i-\tfrac{1}{2})\Delta m_k + m_k^{(\min)}\right)\Delta m_k = 1$$

Table 8. Output format for MPPD.DAT.			
LINE(S)	**PARAMETERS**	**LINE FORMAT**	**DESCRIPTION**
1	*NDIM, MAXNBINS*	(I6, X, I6)	*NDIM* is the number of model parameters. *MAXNBINS* is the maximum value of the array *NBINS*.
2	*NBINS(NDIM)*	(*NDIM* I10)	*NBINS*(*k*) contains the number of histogram bins in the marginal distribution of parameter *k*.
3...*MAXNBINS* + 3	*MPPD(2*NDIM, MAXNBINS)*	(2**NDIM* G15)	*MPPD* is a matrix containing the one-dimensional marginal posterior probability densities for the model parameters. Thus, for each parameter *k*, the x-values of the marginal distribution are listed in *MPPD*(1:*NBINS*(*k*), 2*k*−1) and the y-values are listed in *MPPD*(1:*NBINS*(*k*), 2*k*).

2.5.3 MODELS.DAT

The output file MODELS.DAT contains the entire collection of Gibbs sampled models, accumulated by both independent Gibbs samplers. The format for the MODELS.DAT data file is given in 2.9. The data contained in MODELS.DAT completely characterize the PPD, as estimated by the Bayesian module. Using the data in MODELS.DAT, one can compute various statistical properties of the PPD, including higher-order moments of the marginal distributions (*e.g.*, skewness or kurtosis) and two-dimensional marginal probability distributions. The reader is referred to Dosso's article [1] for details pertaining to these computations.

Table 9. Output format for MODELS.DAT.			
LINE(S)	**PARAMETERS**	**LINE FORMAT**	**DESCRIPTION**
1	*NDIM, NMODELS*	(I6, X, I10)	*NDIM* is the number of parameters in each Gibbs sampled model. *NMODELS* is the number of Gibbs sampled models listed in this file. NOTE: this line is only present *after* the Bayesian PPD module has terminated execution.
2...NMODELS	*MODELS(NDIM, NMODELS)*	(*NDIM* G15)	The collection of Gibbs sampled models. Note that the first model is always the maximum likelihood model.

2.6 Summary of additions/modifications to the ASSA code

The Bayesian PPD module is comprised of a number of different subroutines, each of which is contained in a separate file. All of the files associated with the Bayesian module, along with brief descriptions of their purpose, are listed in Table 10.

Table 10. *Files comprising the Bayesian module.*

FILENAME	SUBROUTINE	FUNCTION
PPD.f	PPD	The main driver routine for the Bayesian PPD module. This routine performs the Gibbs sampling and writes the results to the output files COVM.DAT, MPPD.DAT and MODELS.DAT.
EVALCOST.f	EVALCOST	A wrapper routine for the ASSA COST function. EVALCOST is provided so that modules in the Bayesian code do not have to directly access the ASSA common blocks.
ROTMAT.f	ROTMAT	Computes the parameter rotation matrix, given the model covariance matrix. Uses publicly available EISPACK subroutines (included at the end of the file) for performing the eigenvector decomposition.
PERTROT.f	PERTROT	Randomly perturbs a model vector along a single coordinate axis in a rotated coordinate system.
MATEST.f	MATEST	Performs the Metropolis test for a vector of perturbed model parameters.
UPDATECM.f	UPDATECM	Updates the estimate of the covariance matrix, based on the current collection of Gibbs sampled models.
UPDATEHIST.f	UPDATEHIST	Updates the marginal posterior probability densities of the model parameters based on the current collection of Gibbs sampled models.
CUMDEPS.f	CUMDEPS	Compares two independent estimates of the cumulative distributions for the two independent Gibbs samplers and returns the maximum absolute difference between them.
READ_PPD_DATA.f	READ_PPD_DATA	Inputs the Bayesian specific parts of ASSAIN.DAT. Called by the ASSA routine "ASSA_INPUT".
PPDVARS.inc	—	An include file with common blocks containing the input parameters necessary for running the Bayesian module.

Integrating the Bayesian PPD module into the main ASSA code required several changes to subroutines in the original ASSA program, which are summarized in Table 11. Note that any changes to the original program, pertaining to the Bayesian module, are signed in the code with the initials "AOM" (the initials of the author of the Bayesian module).

Table 11. *Summary of changes to the original ASSA code, pertaining to the Bayesian module.*

FILENAME	LINES	MODIFICATIONS
ASSA.f	53–59	Added PPDVARS.inc to the list of included files.
ASSA.f	157–179	Added code that prints a warning if non-matched-field inversion problems are used in combination with the Bayesian PPD module.
ASSA.f	181–228	Added code that skips directly to the Bayesian module if user parameters *DO_ASSA* and *DO_PPD* are set to false and true respectively.
ASSA.f	656-675	Added code to execute the Bayesian PPD module after the main ASSA minimization.
ASSA_INPUT.f	58–65	Added PPDVARS.inc to the list of included files.
ASSA_INPUT.f	296–318	Added code to input user parameter *DO_PPD* on data line 5 of section 1 of ASSAIN.DAT.
ASSA_INPUT.f	760–767	Added code to call READ_PPD_DATA subroutine.
READ_COSTFN_DATA.f	38–61	Commented out code that triggered an error when trailing input lines appeared in ASSAIN.DAT.

3. References

1. Dosso, Stan E. (2002). "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," *J. Acoust. Soc. Am.*, **111** (1), pp.129-142.

2. Gillard, C.A., Thompson, D.J. and Heard, G.J. (2003). "Estimating Geoacoustic Parameters using Matched-field Inversion Methods," To appear in *IEEE Journal of Oceanic Engineering*.

3. Dosso, S.E. and Nielsen, P.L. (2002). "Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow-water data," *J. Acoust. Soc. Am.*, **111** (1), pp.143-159.

4. Calnan, C. (2002). Adaptive Simplex Simulated Annealing Function Minimization Program. (DRDC Atlantic Contractor Report 2003-033).

5. Dosso, S.E. and Collison, N.E.B. (2000). "Acoustic tracking of a freely drifting sonobuoy field.", *J. Acoust. Soc. Am.*, 111(5), pp. 2166-2177.

6. Dosso, S.E., Fallat, M.R., Sotirin B.J. and Newton, J.L. (1998) "Array element localization for horizontal arrays via Occam's inversion.", *J. Acoust. Soc. Am.*, 104(2), pp. 846-859.

Annex A – Sample Input File for Localization Problem

```
#########################################################################
#
#  Sample ASSA main input file
#
#  Problem 5
#
#  Title

Assartinl test case, 1

#
#  Control codes for:
#     Problem        Model       Cost Function

         5             2              2
#      Specific      Ray Tracing    Specific

#
#  ASSA parameters:
#    T0    Tfact    Ntemp    Npert    NDHS      NAVE     NDIM    True Data

     1.0   0.99     0        5        500       20000    51       .FALSE.

#    NPERTM     PVC     PVC Multiplier  (PVC = Perturbation Value Code)
          5       1              3.
#
#  ASSA Processing Options:
#   CostFn Grid    Do ASSA     Tracking    Screen Prints  Quench Rand.
       .FALSE.       .TRUE.      .TRUE.        .TRUE.        .FALSE.

#  ("Cost Fn Grid" is ignored if "True Data" is .FALSE.)
#  ("Quench Rand." is ignored if "Do ASSA" is .FALSE.)

#  Any parameter selection constraint data go in here

#   NAME LT  3 6
#   NAME LT  3 9
#
#  SSA and DHS convergence values (only if "Do ASSA" = .TRUE.)

      1.0E-2      1.0E-5

#########################################################################
#
#  Problem-specific parameters:
#
# Note that for Problem 5 (or any well behaved problem for which only the
# the downhill simplex, or quenching, part of the minimization is desired)
# set NTEMP above to 0.

# Secondary data input file
  Prob5In.DAT

# Number of receivers
   9

# Receiver positions and uncertainties  - all measured in meters
#      'x'          'y'          'z'       dx        dy        dz
      -500         500          180       10        10        5
         0         500          180       10        10        5
       500         500          180       10        10        5
       500           0          180       10        10        5
         0           0          180       10        10        5
      -500           0          180       10        10        5
      -500        -500          180       10        10        5
         0        -500          180       10        10        5
       500        -500          180       10        10        5

# Number of sources
   8

# Source positions and uncertainties - all measured in meters
```

```
#    'x'        'y'      'z'      dx       dy       dz
      -3000        0      100      50       50        5
       3000        0      100      50       50        5
          0     3000      100      50       50        5
          0    -3000      100      50       50        5
      -3000        0      100      50       50        5
       3000        0      100      50       50        5
          0     3000      100      50       50        5
          0    -3000      100      50       50        5
#####################################################################

#####################################################################
#
#  Allowable limits of the random value for each dimension's parameter
#
#  These data are generic in that every run of ASSA will need them, but
#  their format will be dependant on the problem.

#
#  Number of columns of search parameter limits

     3
#
#  MinLim

   -600  400     99
   -100  400     99
    400  400     99
    400 -100     99
   -100 -100     99
   -600 -100     99
   -600 -600     99
   -100 -600     99
    400 -600     99
  -3001   -1      179
   2999   -1      179
     -1 2999      179
     -1 -3001     179
  -3001   -1      179
   2999   -1      179
     -1 2999      179
     -1 -3001     179


#
#  MaxLim

   -400  600     100
    100  600     100
    600  600     100
    600  100     100
    100  100     100
   -400  100     100
   -400 -400     100
    100 -400     100
    600 -400     100
  -3000    1      180
   3000    1      180
      1 3000      180
      1 -3000     180
  -3000    1      180
   3000    1      180
      1 3000      180
      1 -3000     180

#
#  True (only here if "True Data" = .TRUE.)

##    1.55     1575       0.1228
##    1.5614   1566.4     0.096914
##    1.5614   1566.4     0.096912
##    1.59     1610       0.0322
```

```
##   1.95      1800        0.036
##  10.0         25.0    105.0

************************************************************************
*
*  Model-specific data for Ray Tracing

        0 1497                    * ZI  PROFI
      5.0 1496.5
       10  1494.5
       20  1492
       40  1487
       60  1485.5
       80  1485
      100 1484.5
      120 1484.2
      140 1484
      160 1483.5
      180 1483
      200 1482.2
      250 1482
      300 1481.5


     -1       -1
************************************************************************
*
*  Cost Function-specific
*
```

Annex B – Sample Secondary Input File for Localization Problem

```
1 7
1.7189832
2.0502462
2.3831053
1.6856084
2.0223743
2.3591689
1.7189832
```

Annex C – Sample Input File for Bayesian Module

```
!!!
```

```
  2.200   2100      0.8
  2.200   2100      0.8
  10.0      30.0   91.0

#
#  True (only here if "True Data" = .TRUE.)
#  See user's manual for explanation of this bizarre layout
1.59628 1511.87305 0.02484
0. 1504.45044 0.31628
1.85893 1740.25366 0.48339
0. 1775.67114 0.41094
2.06407 2059.78442 0.10867
4.14358 21.85875 90.09845

#################################################################################
#
#  Model-specific data for RAMGEO
     20.0   30                    # ZS, ZR
      1                           # NDR
      2    800.0                  # NDZ, ZMPLT
   1500.0    5      1      0.0    # C0, NP, NS, RS "all active"

      0.0  200.0                  # RB ZB "ranges active"
   5000.0  200.0
     -1     -1

      0.0 1495.0                  # ZI  PROFI
    160.0 1488.6
     -1     -1

#################################################################################
#
#  Cost Function-specific
#
#  (None for Bartlett)

#################################################################################
#
#  Bayesian PPD input (only if DO_PPD = TRUE)
#
#  EPSILON  MAXSAMP   CASCADE   NCHECK   ESTVAR
   0.10     20000     .FALSE.     5      .TRUE.

#  NF or vF
   4  4

#  NBINS
 20  20  20  20
 20  20  20  20
 20  20  20  20
 20  20  20  20
```

List of symbols/abbreviations/acronyms/initialisms

AEL	Array Element Localization
ASSA	Adaptive Simplex Simulated Annealing
DRDC	Defence Research & Development – Canada
MFP	Matched Field Processing
ML	Maximum Likelihood
PPD	Posterior Probability Density
RMS	Root Mean Square
SNR	Signal to Noise Ratio

Distribution list

NDHQ/DRDKIM 2-2-5 (1)	
DRDC Atlantic (6)	
JASCO Research Ltd. (4)	Suite 2101, 4464 Markham St
	Victoria, BC
	V8Z 7X8
David J. Thomson	733 Lomax Road
	Victoria, BC
	V9C 4A4
Francine Desharnais	DRDC Atlantic
Gordon Ebbeson	DRDC Atlantic
Garry Heard	DRDC Atlantic
H /TD	DRDC Atlantic
H /US	DRDC Atlantic
Nicole Collison	DRDC Atlantic

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

JASCO Research Ltd, Suite 2101, 4464 Markham St, Victoria, BC V8Z 7X8

2. SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable).

UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title).

ASSA inversion model upgrades: implementation of raytrace AEL, and Bayesian posterior probability density distributions

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

Alex MacGillivray and Melanie Austin

5. DATE OF PUBLICATION (month and year of publication of document)	6a. NO. OF PAGES (total containing information Include Annexes, Appendices, etc).	6b. NO. OF REFS (total cited in document)
Jan 2005	32 (approx.)	6

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered).

CONTRACT REPORT

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include address).

Defence R&D Canada – Atlantic
PO Box 1012
Dartmouth, NS, Canada B2Y 3Z7

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant).	9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written).
Project 11cs02	W7707-02-1923

10a ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b OTHER DOCUMENT NOs. (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
	DRDC Atlantic CR 2003-251

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Defence departments and defence contractors; further distribution only as approved
( ) Defence departments and Canadian defence contractors; further distribution only as approved
( ) Government departments and agencies; further distribution only as approved
( ) Defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected).

UNLIMITED

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

The Adaptive Simplex Simulated Annealing (ASSA) Fortran code currently used by DRDC Atlantic to solve a series of geoacoustic inversion problems has been expanded to solve a new type of problem and to provide Bayesian statistical analysis of the model result uncertainties. The new problem type involves the localization of acoustic sources and receivers given relative arrival time data. A new forward model has been added to the program for solving the localization problem using a ray tracing approach. A new cost function, to be minimized in the optimization of the localization problem, has also been added to the program based on the root mean square (rms) time difference of the relative arrival times. The Bayesian analysis is performed using a Gibbs sampling approach to determine *a posteriori* probability density (PPD) distributions, providing estimates of model uncertainties for geoacoustic inversion problems. In order to preserve the general nature of the ASSA code and to allow future program enhancements, the new features have been added in a modular fashion and are pure extensions to the existing code that do not change the main functionality of the original ASSA program.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title).

ASSA
Adaptive simplex simulated annealing
Passive localization
Geo-acoustic inversion
Bayesian analysis
Inversion uncertainties

Defence R&D Canada

Canada's leader in defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

www.drdc-rddc.gc.ca